REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
08-06-2018	Final Report	1-Aug-2016 - 1-Jan-2018

4. TITLE AND SUBTITLE
Final Report: Acquisition of High Performance Computing Instrument for Network Science Research and Education

5a. CONTRACT NUMBER
W911NF-16-1-0416

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
106012

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Nevada - Las Vegas
4505 Maryland Parkway
Box 454019
Las Vegas, NV 89154 -4019

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
68817-NS-REP.17

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Justin Zhan
UU	UU	UU	UU		19b. TELEPHONE NUMBER: 702-895-3681

RPPR Final Report
as of 25-Jun-2018

Agency Code:

Proposal Number: 68817NSREP **Agreement Number: W911NF-16-1-0416**
INVESTIGATOR(S):

Name: Ge Lin Kan
Email: ge.kan@unlv.edu
Phone Number: 7028951625
Principal: N

Name: Haroon Stephen
Email: haroon.stephen@unlv.edu
Phone Number: 7027741463
Principal: N

Name: Justin Zhan
Email: justin.zhan@unlv.edu
Phone Number: 7028953681
Principal: Y

Name: Kazem Taghva
Email: kazem.taghva@unlv.edu
Phone Number: 7028950873
Principal: N

Name: Bing Zhang
Email: zhang@physics.unlv.edu
Phone Number: 7028954050
Principal: N

Organization: **University of Nevada - Las Vegas**
Address: 4505 Maryland Parkway, Las Vegas, NV 891544019
Country: USA
DUNS Number: 098377336 EIN: 886000024

Report Date: 01-Apr-2018 Date Received: 08-Jun-2018
Final Report for Period Beginning 01-Aug-2016 and Ending 01-Jan-2018
Title: Acquisition of High Performance Computing Instrument for Network Science Research and Education
Begin Performance Period: 01-Aug-2016 **End Performance Period:** 01-Jan-2018
Report Term: 0-Other

Submitted By: Justin Zhan Email: justin.zhan@unlv.edu
Phone: (702) 895-3681

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 3 **STEM Participants:** 10

Major Goals: The objective of this proposal is to acquire an integrated high performance computing instrument for big data in network science research and education at University of Nevada Las Vegas.

Accomplishments: Integrated high performance computing instrument for big data in network science research and education has been acquired.

We have used the instrument for various big data research projects including data bridge system project, community detection for large networks project, a deep learning framework for network dynamics prediction project, and more than 50 class projects.

We have also used the instrument for education and course delivery including big data analytics course, deep learning for big data course, and applied big data course.

RPPR Final Report
as of 25-Jun-2018

Training Opportunities: This instrument has strengthened educational training activities by enabling innovative big data courses and promoting our education outreach activity with K-12 teachers and students. A key use of the HPC was to provide training in advanced computing environments and support for faculty and students at University of Nevada Las Vegas (UNLV).

PI is leading a summer program for high/middle school students and teachers. We have 40 students and teachers each summer to attend our 6-week summer program on big data. We provide seminars and demonstrations about bridging network big data from various disciplinary perspectives during this camp to attract students to the ?eld of computing sciences. The demonstrations not only excites high school students but also educates them about network big data problems facing the digital age.

Results Dissemination: We gave presentations at various conferences such as International Conference on Big Data. Our presentations aimed at scholars whose research and education interests are on big data analytics. We also published our research results in peer-reviewed journals.

Honors and Awards: Congressional Recognition Award, US Congresswoman Jack Rosen, Member of Congress, Third District, Nevada, July 2017.

The United States Army Mentorship Award for Research & Engineering Apprenticeship Program, July 2017.

Distinguished Researcher Award, College of Engineering, University of Nevada Las Vegas, April 2017.

Outstanding Researcher Award, Department of Computer Science, University of Nevada Las Vegas, April 2017.

Faculty Mentor Award, Center for Academic Enrichment and Outreach, University of Nevada Las Vegas, November 2016.

Faculty Fellow, Air Force Research Laboratory, May-August 2016.

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: PD/PI
Participant: Justin Zhan
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: Kazem Tagva
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: Ge Lin Kan
Person Months Worked: 1.00 **Funding Support:**

RPPR Final Report
as of 25-Jun-2018

Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: Bing Zhang
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co-Investigator
Participant: Haroon Stephen
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Access
Publication Identifier Type: DOI Publication Identifier: 10.1109/ACCESS.2018.2797048
Volume: 6 Issue: First Page #: 7872
Date Submitted: 6/8/18 12:00AM Date Published:
Publication Location:
Article Title: Finding Top- k Dominance on Incomplete Big Data Using MapReduce Framework
Authors: Payam Ezatpoor, Justin Zhan, Jimmy Ming-Tai Wu, Carter Chiu
Keywords: Top-k dominance, incomplete data, bigdata, mapreduce, hadoop, dominance relationship, query processing.
Abstract: Incomplete data is one major kind of multi-dimensional dataset that has random-distributed missing nodes in its dimensions. It is very difficult to retrieve information from this type of dataset when it becomes large. Finding top-k dominant values in this type of dataset is a challenging procedure. Some algorithms are present to enhance this process, but most are efficient only when dealing with small incomplete data. This paper proposes MapReduced Enhanced Bitmap Index Guided Algorithm (MRBIG) for dealing with the aforementioned issues. MRBIG uses the MapReduce framework to enhance the performance of applying top-k dominance queries on large incomplete datasets. The proposed approach uses the MapReduce parallel computing approach involving multiple computing nodes. The framework separates the tasks between several computing nodes to independently and simultaneously work to find the result.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 25-Jun-2018

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Big Data
Publication Identifier Type: DOI Publication Identifier: 10.1109/TBDATA.2016.2628725
Volume: 3 Issue: 3 First Page #: 276
Date Submitted: 6/8/18 12:00AM Date Published: 9/1/17 7:00AM
Publication Location:
Article Title: A Framework for Community Detection in Large Networks Using Game-Theoretic Modeling
Authors: Pravin Chopade, Justin Zhan
Keywords: Game theory, large-scale networks, community detection, adjacency, Laplacian, modularity, similarity, clustering
Abstract: This paper presents a new game-theoretic approach towards community detection in large-scale complex networks based on modified modularity; this method was developed based on modified adjacency, modified Laplacian matrices and neighborhood similarity.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Access
Publication Identifier Type: DOI Publication Identifier: 10.1109/ACCESS.2017.2771448
Volume: 5 Issue: First Page #: 27354
Date Submitted: 6/8/18 12:00AM Date Published:
Publication Location:
Article Title: A Novel Weak Estimator For Dynamic Systems
Authors: Moinak Bhaduri, Justin Zhan, Carter Chiu
Keywords: Stochastic learning weak estimators, dynamic systems, continuous time Markov chain, countable state space, non-stationarity, classification, simulation
Abstract: In this paper, we propose a novel approach for classifying incoming continuous data under a non-stationary environment. A class of estimators termed stochastic learning weak estimators has been generalized to include continuous time sampling and countable state categories. The method is founded on non-stationary Markov chain techniques and is useful in diverse applications, such as consumer behavior analysis, e-mail spam classification, or understanding drug effectiveness. In terms of tracking the true state probabilities, these weak estimators consistently outperform traditional competitors such as maximum likelihood estimates. Only one user defined parameter is necessary and the method is free of subjective ‘‘moving window’’ type algorithms. We have conducted extensive simulations and real data analyses for classification purposes.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Access
Publication Identifier Type: DOI Publication Identifier: 10.1109/ACCESS.2017.2773038
Volume: 5 Issue: First Page #: 26364
Date Submitted: 6/8/18 12:00AM Date Published:
Publication Location:
Article Title: Toward Efficient Hub-Less Real Time Personalized PageRank
Authors: Matin Pirouz, Justin Zhan
Keywords: Graph theory, hub nodes, PageRank, personalized PageRank
Abstract: In the era of big data, reduced models capable of reducing big data graph to estimate personalized PageRank are limited. Personalized PageRank is a page rank calculation where random jumps are only allowed to a subset of start nodes. The resources of current process of calculation of personalized PageRank are highly prohibitive, thus in this paper we propose a novel fast accurate and less resource intensive algorithm to the personalized PageRank problem. FAST Personalized PageRank is utilized to find the target node set. Using the mentioned target set, the algorithm gives an estimation of the closeness of any pair of nodes in the graph. As the time taken by the estimation of personalized PageRank is directly proportional to the network size, in this paper a node reduction method is used to prune the graph.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 25-Jun-2018

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Access
Publication Identifier Type: DOI Publication Identifier: 10.1109/ACCESS.2017.2735378
Volume: 5 Issue: First Page #: 15883
Date Submitted: 6/8/18 12:00AM Date Published:
Publication Location:
Article Title: A Novel Online and Non-Parametric Approach for Drift Detection in Big Data
Authors: Moinak Bhaduri, Justin Zhan, Carter Chiu, Felix Zhan
Keywords: Change point detection, non-parametric methods, Hoeffding's inequality, Bernstein's inequality, big data, anomaly detection
Abstract: A sizable amount of current literature on online drift detection tools thrive on unrealistic parametric strictures such as normality or on non-parametric methods whose power performance is questionable. Using minimal realistic assumptions such as unimodality, we have strived to proffer an alternative, through a novel application of Bernstein's inequality. Simulations from such parametric densities as Beta and Logitnormal as well as real-data analyses demonstrate this new method's superiority over similar techniques relying on bounds, such as Hoeffding's. Improvements are apparent in terms of higher power, efficient sample sizes, and sensitivity to parameter values.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Access
Publication Identifier Type: DOI Publication Identifier: 10.1109/ACCESS.2017.2726078
Volume: 5 Issue: First Page #: 13689
Date Submitted: 6/8/18 12:00AM Date Published:
Publication Location:
Article Title: Uncovering Suspicious Activity From Partially Paired and Incomplete Multimodal Data
Authors: Carter Chiu, Justin Zhan, Felix Zhan
Keywords: Suspicious activity, multimodal data, partially paired data, incomplete data
Abstract: Multimodal data can be used to gain additional perspective on a phenomenon. For applications, such as security and the detection of suspicious activity, the need to aggregate and analyze data from multiple modes is vital. Recent research in suspicious behavior detection has introduced methods for identifying and scoring dense blocks in multivariate tensors, which are consistent indicators of suspicious activity. None yet, however, have proposed a method for the merging and analysis of multiple modes of data for suspicious behavior, especially when the set of items described in each data set do not match—that is, the data is partially paired—which is common when data sets originate from different sources. Neither has a method been described for dealing with the similar case of incomplete data. This paper introduces a technique for multimodal data analysis for suspicious activity detection when the data are only partially paired and/or incomplete. The method is applied to synthetic and r
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 25-Jun-2018

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Access
Publication Identifier Type: DOI Publication Identifier: 10.1109/ACCESS.2017.2723838
Volume: 5 Issue: First Page #: 13046
Date Submitted: 6/8/18 12:00AM Date Published:
Publication Location:
Article Title: Using Proxies for Node Immunization Identification on Large Graphs
Authors: Raymond Ahn, Justin Zhan
Keywords: Immunization, linear time, proxies, tree-graph
Abstract: Given a large graph, like a social network, which k nodes should be immunized (or removed) to make the network safe from the spread of a virus? This is the node immunization problem. One of the classical methods, inspired by immunology, in analyzing this problem relies on the calculation of the largest eigenvalue before and after immunization in order to create the largest difference in eigenvalue. We propose a method that does not rely on a costly calculation of eigenvalues; instead, we rely on the notion of proxies and deterministic routing areas in order to find such nodes to immunize. We show that our results are consistent with the notion of vulnerability and produces equivalent results when compared with the existing algorithms. Furthermore, experimental results show that when a virus is not allowed to die out (controlled by the strength of the virus), our algorithm ensures that more nodes are safe from infection.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Access
Publication Identifier Type: DOI Publication Identifier: 10.1109/ACCESS.2017.2702281
Volume: 5 Issue: First Page #: 10024
Date Submitted: 6/8/18 12:00AM Date Published:
Publication Location:
Article Title: Ant Colony System Sanitization Approach to Hiding Sensitive Itemsets
Authors: Jimmy Ming-Tai Wu, Justin Zhan, Jerry Chun-Wei Lin
Keywords: Privacy-preserving data mining, evolutionary algorithm, sensitive itemsets, ant colony system
Abstract: In this paper, an ant colony system (ACS)-based algorithm called ACS2DT is proposed to decrease side effects and enhance the performance of the sanitization process. Each ant in the population will build a tour for each iteration and each tour indicates the deleted transactions in the original database. The proposed algorithm introduces a useful heuristic function to conduct each ant to select a suitable edge (transaction) for the current situation and also designs several termination conditions to stop the sanitization processes. The proposed heuristic function applies the pre-large concept to monitor side effects and calculates the degree of hiding information to adjust the selecting policy for deleted transactions. The experimental results show that the proposed ACS2DT algorithm performs better than the Greedy algorithm and other two evolutionary algorithms in terms of runtime, fail to be hidden, not to be hidden, not to be generated and database similarity on both real-world and sy
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 25-Jun-2018

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Big Data
Publication Identifier Type: DOI Publication Identifier: 10.1186/s40537-017-0095-2
Volume: 4 Issue: 1 First Page #:
Date Submitted: 6/8/18 12:00AM Date Published: 10/1/17 12:00AM
Publication Location:
Article Title: Using deep learning for short text understanding
Authors: Justin Zhan, Binay Dahal
Keywords: Short text classifcation, Semantic enrichment, Deep neural network
Abstract: Classifying short texts to one category or clustering semantically related texts is challenging, and the importance of both is growing due to the rise of microblogging platforms, digital news feeds, and the like. We propose to address this issue using semantic enrichment. This is accomplished by taking the nouns, and verbs used in the short texts and generating the concepts and co-occurring words with the help of those terms. The nouns are used to generate concepts within the given short text, whereas the verbs are used to prune the ambiguous context (if any) present in the text. The enriched text then goes through a deep neural network to produce a prediction label for that short text representing it's category.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Big Data
Publication Identifier Type: DOI Publication Identifier: 10.1186/s40537-017-0076-5
Volume: 4 Issue: 1 First Page #:
Date Submitted: 6/8/18 12:00AM Date Published: 5/1/17 7:00AM
Publication Location:
Article Title: Identification of top-K nodes in large networks using Katz centrality
Authors: Justin Zhan, Sweta Gurung, Sai Phani Krishna Parsa
Keywords: Top-K nodes, Katz centrality, Social networks
Abstract: In this paper, we propose a novel method for identifying top-K viral information propagators from a reduced search space. Our algorithm computes the Katz centrality and Local average centrality values of each node and tests the values against two threshold (constraints) values. Only those nodes, which satisfy these constraints, form the search space for top-K propagators.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Big Data
Publication Identifier Type: DOI Publication Identifier: 10.1186/s40537-016-0061-4
Volume: 4 Issue: 1 First Page #:
Date Submitted: 6/8/18 12:00AM Date Published: 1/1/17 8:00AM
Publication Location:
Article Title: Vaccination allocation in large dynamic networks
Authors: Justin Zhan, Timothy Rafalski, Gennady Stashkevich, Edward Verenich
Keywords: Dynamic networks, Immunization, Dominance tree
Abstract: Network infections that are already in progress cause challenges to those officers trying to preserve those nodes not yet infected. Static solutions can take advantage of global knowledge of the network to produce quick and approximate answers for those members who should be vaccinated. In dynamic situations however, small changes can severely alter those static solutions making them irrelevant. Yet in dynamic situations it can not be known with certainty which small changes will affect the solution and those that will not. Computational resources are wasted recalculating a global solution for the entire network, when a local recalculation may be enough. This paper presents a dynamic node vaccination solution that seeks to take advantage of these local recalculations
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 25-Jun-2018

DISSERTATIONS:

Publication Type: Thesis or Dissertation
Institution: University of Nevada Las Vegas
Date Received: 08-Jun-2018 Completion Date: 5/31/17 8:08PM
Title: FINDING TOP-K DOMINANCE ON INCOMPLETE BIG DATA USING MAP-REDUCE FRAMEWORK
Authors: Payam Ezatpoor
Acknowledged Federal Support: **N**

Publication Type: Thesis or Dissertation
Institution: University of Nevada Las Vegas
Date Received: 08-Jun-2018 Completion Date: 5/1/18 7:00AM
Title: ADVANCING COMMUNITY DETECTION USING KEYWORD ATTRIBUTE SEARCH
Authors: Sanket Chobe
Acknowledged Federal Support: **N**

Publication Type: Thesis or Dissertation
Institution: University of Nevada Las Vegas
Date Received: 08-Jun-2018 Completion Date: 5/31/17 7:00AM
Title: HIGH UTILITY ITEMSETS IDENTIFICATION IN BIG DATA
Authors: Ashish Tamrakar
Acknowledged Federal Support: **N**

Nothing to report in the uploaded pdf